

<https://helda.helsinki.fi>

Delineating reef fish trophic guilds with global gut content data synthesis and phylogeny

Parravicini, Valeriano

2020-12

Parravicini , V , Casey , J M , Schiettekatte , N M D , Brandl , S J , Pozas-Schacre , C , Carlot , J , Edgar , G J , Graham , N A J , Harmelin-Vivien , M , Kulbicki , M , Strona , G & Stuart-Smith , R D 2020 , ' Delineating reef fish trophic guilds with global gut content data synthesis and phylogeny ' , PLoS Biology , vol. 18 , no. 12 , 3000702 . <https://doi.org/10.1371/journal.pbio.3000702>

<http://hdl.handle.net/10138/325113>

<https://doi.org/10.1371/journal.pbio.3000702>

cc_by

publishedVersion

Downloaded from Helda, University of Helsinki institutional repository.

This is an electronic reprint of the original article.

This reprint may differ from the original in pagination and typographic detail.

Please cite the original version.

METHODS AND RESOURCES

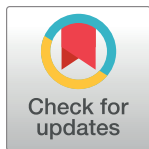
Delineating reef fish trophic guilds with global gut content data synthesis and phylogeny

Valeriano Parravicini^{1,2,*}, Jordan M. Casey^{1,2,3}, Nina M. D. Schiettekatte^{1,2}, Simon J. Brandl^{1,2,3,4}, Chloé Pozas-Schacre^{1,2}, Jérémy Carlot^{1,2}, Graham J. Edgar⁵, Nicholas A. J. Graham⁶, Mireille Harmelin-Vivien⁷, Michel Kulbicki⁸, Giovanni Strona⁹, Rick D. Stuart-Smith⁴

1 PSL Université Paris: EPHE-UPVD-CNRS, USR 3278 CRIOBE, Université de Perpignan, Perpignan, France, **2** Laboratoire d'Excellence "CORAIL," Perpignan, France, **3** Department of Marine Science, University of Texas at Austin, Marine Science Institute, Port Aransas, Texas, United States of America, **4** Centre for the Synthesis and Analysis of Biodiversity (CESAB), Institut Bouisson Bertrand, Montpellier, France, **5** Institute for Marine and Antarctic Studies, University of Tasmania, Hobart, Tasmania, Australia, **6** Lancaster Environment Centre, Lancaster University, Lancaster, United Kingdom, **7** Aix-Marseille Université, Institut Méditerranéen d'Océanologie, CNRS/INSU, Marseille, France, **8** UMR Entropie, LabEx Corail, IRD, Université de Perpignan, Perpignan, France, **9** University of Helsinki, Department of Bioscience, Helsinki, Finland

* These authors contributed equally to this work.

* valeriano.parravicini@ephe.psl.eu



OPEN ACCESS

Citation: Parravicini V, Casey JM, Schiettekatte NMD, Brandl SJ, Pozas-Schacre C, Carlot J, et al. (2020) Delineating reef fish trophic guilds with global gut content data synthesis and phylogeny. *PLoS Biol* 18(12): e3000702. <https://doi.org/10.1371/journal.pbio.3000702>

Academic Editor: Pedro Jordano, Estacion Biologica de Doñana CSIC, SPAIN

Received: February 13, 2020

Accepted: December 3, 2020

Published: December 28, 2020

Copyright: © 2020 Parravicini et al. This is an open access article distributed under the terms of the [Creative Commons Attribution License](https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original author and source are credited.

Data Availability Statement: All methods are described within the paper and its [Supporting Information](#) files. The data and the R code are available at the GitHub repository https://github.com/valerianoparravicini/Trophic_Fish_2020.

Funding: This research was funded by the BNP Paribas Foundation (Reef Services Project) and the French National Agency for Scientific Research (ANR; REEFLEX Project; ANR-17-CE32-0006). This research is product of the SCORE-REEF group funded by the Centre de Synthèse et d'Analyse sur

Abstract

Understanding species' roles in food webs requires an accurate assessment of their trophic niche. However, it is challenging to delineate potential trophic interactions across an ecosystem, and a paucity of empirical information often leads to inconsistent definitions of trophic guilds based on expert opinion, especially when applied to hyperdiverse ecosystems. Using coral reef fishes as a model group, we show that experts disagree on the assignment of broad trophic guilds for more than 20% of species, which hampers comparability across studies. Here, we propose a quantitative, unbiased, and reproducible approach to define trophic guilds and apply recent advances in machine learning to predict probabilities of pairwise trophic interactions with high accuracy. We synthesize data from community-wide gut content analyses of tropical coral reef fishes worldwide, resulting in diet information from 13,961 individuals belonging to 615 reef fish. We then use network analysis to identify 8 trophic guilds and Bayesian phylogenetic modeling to show that trophic guilds can be predicted based on phylogeny and maximum body size. Finally, we use machine learning to test whether pairwise trophic interactions can be predicted with accuracy. Our models achieved a misclassification error of less than 5%, indicating that our approach results in a quantitative and reproducible trophic categorization scheme, as well as high-resolution probabilities of trophic interactions. By applying our framework to the most diverse vertebrate consumer group, we show that it can be applied to other organismal groups to advance reproducibility in trait-based ecology. Our work thus provides a viable approach to account for the complexity of predator–prey interactions in highly diverse ecosystems.

la Biodiversité (CESAB) of the Foundation pour la Recherche sur la Biodiversité (FRB) and the Agence Nationale de la Biodiversité (AFB). VP was supported by the Institut Universitaire de France (IUF), JMC was supported by a Make Our Planet Great Again Postdoctoral Grant (mopga-pdf-0000000144) and JC was supported by the French Polynesian Government (RisqueRecif project). The funders had no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript.

Competing interests: The authors have declared that no competing interests exist.

Abbreviations: AUC, area under the receiver operating characteristic curve; HMD, herbivores, microvores, and detritivores; TSS, true skills statistics.

Introduction

A fundamental goal in ecology is to understand the mechanisms behind the maintenance of biodiversity and ecosystem functioning [1,2]. Understanding the ecological niches of species and their role in ecosystems is central to this endeavor [3,4]. In fact, the degree of niche overlap among species can be a major determinant of relationships among species richness [5], ecosystem productivity [6–8], and vulnerability [9] since limited functional redundancy can make ecosystems more prone to lose entire energetic pathways [10–12]. With growing threats to flora and fauna worldwide, the need to quantify the impact of biodiversity loss has amplified the use of functional groups, which group species (and life history stages) that share common ecological characteristics and are often defined by coarse, categorical descriptors of species traits [13–16].

Natural systems are inherently complex, with almost innumerable, non-random linkages across an intricate network of ecological interactions [17]. Accounting for such complexity is critical to define energetic pathways and, ultimately, ecosystem functioning [18]. However, our understanding of even basic predator–prey interactions is limited for many ecosystems, and expert opinion does not adequately fill this knowledge gap [19]. To overcome this limitation, scientists have developed methods to infer the probability of ecological interactions based on species' evolutionary history and ecological traits [20–23]. However, predicting trophic interactions across the entire spectrum of potential predator–prey interactions often remains unresolved in hyperdiverse ecosystems. In these cases, categorical traits are frequently used as proxy of both ecosystem functioning and trophic structure [24].

Delineating the ecological niche with discrete categories has several operational advantages. First, grouping species into categories helps decompose highly complex ecosystems into comprehensible units, while traditional taxonomic analyses may be difficult to interpret. Second, ecological predictions tied to species are restricted to the geographic range of the species, whereas predictions of functional groups can be globally comparable. Third, the use of functional groups enables the quantification of functional metrics (e.g., functional richness and functional redundancy) from a standard community data matrix without complex manipulative experiments [25–27]. The promise of “user-friendly” metrics for functional ecology has encouraged the employment of trait-based data in community ecology; even with a paucity of empirical information, it is often assumed that experts can achieve accurate descriptions of the ecological niche of species [25,28,29].

Coral reefs, one of the most diverse ecosystems on Earth, have inspired a plethora of studies that assess ecosystem functioning. However, only few studies have attempted to categorize fluxes on a continuous gradient across an entire food web [30], and most studies use expert opinion to define simple functional groups. Indeed, recent efforts have compiled trait-based datasets for 2 major components of this ecosystem: corals and fishes [31,32]. For some traits, such as maximum body size in fishes, the compilation process is simple and accurate because unidimensional, quantitative data (e.g., maximum total length) are compiled in publicly accessible databases; however, when it comes to species' diet or behavior, obtaining consensual data is much more difficult. For example, dietary data are multidimensional (i.e., various prey items can be recorded across individuals), influenced by ontogenetic and spatio-temporal variables (i.e., life history, time, and location can incur dietary shifts), and prone to methodological differences and thus observer bias. Therefore, while some exceptions exist [30,33], our capacity to define coral reef trophic interactions still largely depends on discrete trophic categories defined by expert opinion [27].

Although experts sometimes agree on relevant traits to define trophic categories, there is often an implicit disagreement. Across the coral reef literature, the number and resolution of

reef fish trophic guilds substantially differs. Studies commonly define 3 [34] to 8 [35] trophic guilds, with particular ambivalence on the resolution at which to define herbivores and invertivores [36–39]. Among all trait classification schemes for reef fishes, only a few are openly accessible (e.g., [39,40]). Consequently, different research groups tend to employ proprietary classifications, with little possibility to cross-check and compare assigned traits. The classification of species into functional groups has advantages for our understanding of ecological patterns [42,43]. However, the lack of agreement and the limited transparency of trait-based datasets can conjure skepticism and inhibit the emergence of general patterns.

Here, we quantify expert agreement in the definition of coral reef fish trophic guilds and propose a novel, quantitative framework to delineate trophic guilds. Moreover, we test whether machine learning allows us to go beyond the definition of discrete categories, accurately predicting individual trophic interactions in hyperdiverse ecosystems. We compiled all quantitative, community-wide dietary analyses from several locations across the Indo-Pacific and the Caribbean. Then, we used network analysis to quantitatively define modules that correspond to trophic guilds and machine learning to infer pairwise trophic interactions. We then examined phylogenetic niche conservatism between species to predict trophic guilds and probabilities of pairwise trophic interactions for the global pool of coral reef fishes. Our framework is fully reproducible and can be extended and updated as new data become available.

Materials and methods

Assessment of expert agreement

We systematically searched Google Scholar, including papers since 2000, using the following keywords: “coral reefs” AND “reef fish” AND (“fish community” OR “fish assemblage”) AND “diet” AND (“functional group” OR “functional trait” OR “functional entity” OR “trophic guild” OR “trophic group”). The search yielded 856 papers, which were individually assessed. We only kept studies performed at the community level that targeted all trophic levels. Most studies were excluded because they only included specific families or groups, or the data were not provided with the publication. When the data were not provided with the publications, we contacted authors with trophic classifications widely used across the literature. We often found redundant results, with groups publishing several papers using the same classification scheme. In those cases, only the most recent reference was retained. Of the 856 papers, 163 papers were inaccessible (i.e., non-English language and/or data inaccessibility despite contacting the first author). Thus, 182 studies met the criteria of our initial assessment, which ultimately yielded 33 papers with independent trophic classifications (S1 Table).

The classifications were not uniform in terms of the number and nature of trophic guilds. In order to compare trophic guilds across publications, we first standardized the schemes by converting the original trophic categories into 5 broad trophic guilds: “herbivores and detritivores,” “invertivores,” “omnivores,” “planktivores,” and “piscivores.” All classification schemes could be attributed to these categories with the exception of 8 papers that did not include either “omnivores” or “piscivores” as a category. In these cases, the comparison was only made across the 4 comparable guilds.

In order to assess expert agreement, we compared each possible pair of classifications that shared at least 50 species, generated a confusion matrix (also known as an error matrix; [50]), and measured agreement as the proportion of species with matching trophic guild assignments. We then calculated the average agreement between classification pairs for each trophic guild. Simplifying categories into 5 comparable, broad trophic guilds therefore reduced the number of trophic categories and naturally inflated agreement among authors; thus, our estimates of author agreement are conservative.

Data collection on fish gut contents

To provide a quantitative definition of trophic guilds for reef fishes, we collected gut content data across the literature at the individual or species level for Elasmobranchii (i.e., cartilaginous fishes) and Actinopterygii (i.e., ray-finned fishes). We obtained dietary information from 5 published works: Hiatt and Strasburg (1960) for the Marshall Islands [51], Randall (1967) for Puerto Rico and the Virgin Islands [52], Hobson (1974) for Hawaii [53], Harmelin-Vivien (1979) for Madagascar [54], and Sano and colleagues (1984) for Okinawa [55]. In addition, we provide hitherto unpublished data on the gut contents of 3,015 individuals of 111 species collected in New Caledonia from 1984 to 2000.

All dietary information was based on visual gut content analysis that reported prey ingestion as a volumetric percentage or frequency. The data were standardized and analyzed as proportions. To our knowledge, the compiled dataset represents the first detailed synthesis of community-wide visual gut content analyses to infer the structure of coral reef food webs across ocean basins. A total of 13,961 non-empty fish guts belonging to 615 species were analyzed, and more than 1,200 different prey items were described across the original datasets.

First, fish species and family names were taxonomically verified and corrected with the R package *rfishbase* [56]. Only species with at least 10 non-empty guts were kept for further analysis. The taxonomic classification of each prey item was then obtained, and all non-informative and redundant items were discarded (e.g., unidentified fragments; “crustacea fragments” when co-occurring with an item already identified to a lower taxonomic level such as “shrimp”). Prey identification was highly heterogeneous across the 6 datasets, differing in taxonomic level and the use of common or scientific names (e.g., crabs versus *Brachyura*). In order to make the 6 datasets comparable, prey items were grouped into 38 ecologically informative prey groups (S2 Table). Items were generally assigned to groups corresponding to their phylum or class. Due to the high diversity and detailed descriptions of crustaceans, they were assigned to the level of order or superorder. Most groups follow official taxonomic classifications except for “detritus,” “inorganic,” and “zooplankton.” In the West Indies dataset [52], items labelled as “Algae & Detritus” were assigned to both of the categories “detritus” and “benthic autotroph,” and the percentage was equally divided in 2. The category “zooplankton” includes all eggs and larvae regardless of taxonomy.

Definition of trophic guilds with network analysis

Of the 615 species with dietary information, 516 were present in only 1 location, 66 were collected in 2 locations, 25 in 3 locations, 7 in 4 locations, and only 1 across 5 locations. We tested whether there was a strong dietary difference in species present in more than 1 location by creating a quantitative bipartite network [57] where fish species at each location were linked to the 38 prey groups. This network was weighted so that edge weights represent the proportional contribution of each prey group to the diet of a species at a given location.

In order to identify network modules that correspond to reef fish trophic guilds and their ingested prey, we used the maximization of the weighted network modularity based on weighted bipartite networks [58]. Due to the high occurrence of accidental predation in reef fishes, we used weighted networks to define modules so that rare or accidental prey would not drive the definition of trophic guilds.

Since the modularity maximization algorithm has an initial random step, it may converge to different (although similar) suboptimal solutions each time the analysis is performed, which is common across several optimization algorithms, such as simulated annealing [59]. To guarantee reproducibility and reduce the risk of basing our analysis on an outlier, we performed the modularity maximization 500 times and retained the medoid solution, which was

identified as the solution with the highest similarity to the other 499 modules. Similarity between classifications was assessed as the variation of information, which is an accepted metric to compare multiple clustering results [60]. Overall, 68% of the species found in more than 1 location belonged to the same module. Therefore, we considered the regional effect to be minor and performed the analysis on the global network, ignoring regional variability and increasing the number of individuals per species.

Phylogenetic conservatism of trophic guilds

We extracted the phylogenetic position of the 615 species used for the definition of trophic guilds through the Fish Tree of Life [61]. A total of 603 out of 615 species were available in the Fish Tree of Life, but only 535 species had verified phylogenetic information. For the taxa available in the Fish Tree of Life without verified phylogenetic information, we retrieved the pseudo-posterior distribution of 100 synthetic stochastically resolved phylogenies where missing taxa were placed according to taxonomy using the function *fishtree_complete_phylogeny()* in the R package *fishtree* [62].

We quantified the phylogenetic conservatism of trophic guilds by calculating the phylogenetic statistic δ , which uses a Bayesian approach for discrete variables [63]. The δ statistic can be arbitrarily large with a high level of variation, depending on the number of species and trait levels. To evaluate the significance of the δ statistic, we applied a bootstrapping approach where we quantified δ 100 times after randomly shuffling the trait values.

We then fitted a multinomial phylogenetic regression to predict fish trophic guild according to phylogeny and body size with the R package *brms* [64]. We used a multinomial logit link function. As such, the probability of a particular trophic guild is computed as follows:

$$Pr(k|mu_1, mu_2, \dots, mu_k) = \frac{mu_k}{\sum_1^k \exp(mu_i)} \quad (1)$$

with mu_k defined as

$$mu_1 = 0, mu_{k|2:s} = \beta_{0k} + \beta_{1k} \log(\text{sizemax}) + \gamma_{0\text{phy} \times k}, \quad (2)$$

where β_{0k} is the category-specific fixed-effect intercept, β_{1k} is the slope for the natural transformed maximum body size for each category k , and $\gamma_{0\text{phy} \times k}$ is the matrix of random effect coefficients that account for intercept variation based on relatedness as described by the phylogeny for each diet category k . We used uninformative priors and ran the model for 3 chains, each with 6,000 iterations and a warm-up of 1,000 iterations. We visualized the fitted probabilities for each trophic guild with a phylogenetic tree, including the 535 species with verified phylogenetic positions using the R package *ggtree* [63]. Next, we used our model to predict the most likely trophic guild for the global pool of reef fish species. For the extrapolation, we selected all species within reef fish families with more than 1 representative species (but we also included *Zanclus cornutus*, which is the only species in the family Zanclidae), which resulted in 50 families. Further, we only selected species with a maximum length greater than 3 cm, which was the maximum size of the smallest fish in our compiled database. This selection process resulted in a list of 4,554 reef fish species.

Currently, no streamlined method exists to predict traits for new species from a phylogenetic regression model. We circumvented this issue by extracting draws of the phylogenetic effect ($\gamma_{0\text{phy} \times k}$) for each species included in the model. We subsequently predicted the phylogenetic effects for missing species with the help of the function *phyEstimate* from the R package *picante* [65]. This function uses phylogenetic ancestral state estimation to infer trait values for new species on a phylogenetic tree by re-rooting the tree to the parent edge to predict the node [66]. We repeated this inference across 2,000 draws. Per draw, we randomly sampled 1 of the

100 trees. Then, we predicted the probability of each species to be assigned to each diet category by combining the predicted phylogenetic effects with the global intercept and slopes for maximum body size for each draw. Finally, we summarized all diet category probabilities per species by taking the mean and standard deviation across all 2,000 draws.

We quantified the total standard deviation (i.e., the square root of the quadratic sum of the standard deviations in each category) and the negentropy value, a measure of certainty calculated by subtracting 1 from the entropy value (i.e., uncertainty). Thus, the negentropy value lies between 0 and 1, and the higher the value, the higher the certainty for trophic guild assignment (i.e., if a given species has a high probability of assignment to a dietary category, the negentropy value will be high).

Finally, we conducted a cross validation to validate our extrapolation of trophic guilds to the global pool of fish species. Specifically, we repeated the extrapolation approach (as described above) 535 times, each time leaving out 1 species and predicting the trophic guild of that species. We then compared this prediction to the original assigned trophic guild and calculated the accuracy of each of the 8 trophic guild predictions.

Prediction of trophic interactions with machine learning

To complement the assignment of discrete trophic guilds, we also modeled pairwise trophic interactions. In accordance with previous studies that infer trophic interactions by matching species traits or phylogenetic position [19,21,23,67,68], we predicted the probability of pairwise trophic interactions between the 535 reef fish species and the 38 prey categories in our dataset. Building on Laigle and colleagues [21], we developed a new machine learning approach to extract the reef fish phylogenetic tree from the Fish Tree of Life [61] and obtain phylogenetic eigenvector maps for each species, which were used as explanatory variables in our models [69]. We then predicted the probability of trophic interactions between fish species and prey categories based on phylogenetic position and maximum body size. Specifically, using the R package *h2o* [70], we employed an ensemble modeling approach based on 3 models calibrated with 10-fold cross validation: extreme gradient boosting [71], boosted regression trees [72], and random forest [73]. A cross-validated general linear model was used as a super-learner to create an optimal weighted average (i.e., an ensemble) of the predictions from the 3 models. The 3 models were implemented using 2,000 regression trees and default settings to reduce overfitting. Model performance was assessed using the area under the receiver operating characteristic curve (AUC) and true skills statistics (TSS) [74].

In addition to applying this analysis to our dataset, we also tested whether this technique could reliably predict pairwise trophic interactions for new species and locations. To this aim, we calibrated the models with only 5 locations, excluding the dataset from New Caledonia. We then used the New Caledonia dataset to assess model performance. As detailed above, after cross validation, we used our model to predict probabilities of pairwise trophic interactions between the 4,554 reef fish species and the 38 prey categories.

Results

Assessment of expert agreement

We evaluated expert agreement among 33 distinct and independent trophic guild classifications by comparing the classification schemes in pairs. Considering the broadness of the expert-assigned categories, we found low agreement. The median agreement between pairs, expressed as the proportion of species with matching trophic group assignments, was 78% (Fig 1). For approximately 50% of the pairwise comparisons, at least a quarter of the species were attributed to different trophic groups. In the most severe disagreement, the proportion of

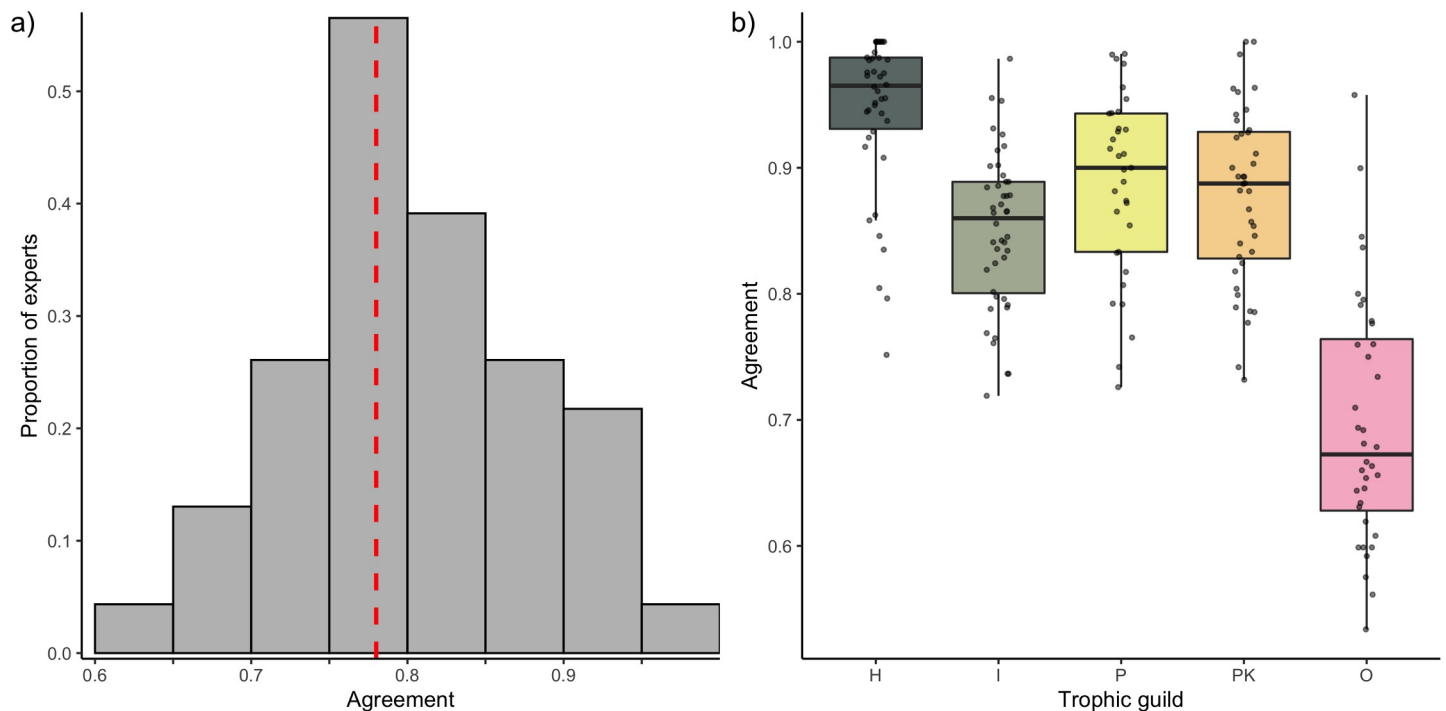


Fig 1. Expert agreement on trophic guild assignment. (A) The distribution of the agreement (i.e., proportion of species assigned to the same trophic category) across the 32 comparisons between pairs of experts. The red dotted line represents the median. (B) Agreement between pairs of experts by trophic category. The data underlying this figure may be found in https://github.com/valerianoparravicini/Trophic_Fish_2020. H, herbivores and detritivores; I, invertivores; O, omnivores; P, piscivores; PK, planktivores.

<https://doi.org/10.1371/journal.pbio.3000702.g001>

mismatched assignments reached 38%. In addition, expert agreement differed depending on the trophic group. Expert disagreement on the classification of “herbivores and detritivores” was low, with an average expert agreement of 95% and pairs of expert disagreement only reaching 20% (Fig 1B). In contrast, “omnivores” showed the highest mismatch, with experts agreeing on an average of only 70% of the species and peaks of disagreement between expert pairs reaching 47% (Fig 1B).

Expert agreement was variable and often homogeneously distributed around the mean for all the trophic categories. Therefore, the high agreement between a few combinations of experts did not necessarily exclude peaks of disagreement (Fig 1B). The analysis of individual confusion matrices between pairs of experts revealed high heterogeneity (Fig 2). For example, Moraes and Bellwood were generally in agreement with Mouillot and colleagues [36] (across 89% of the 515 species in common), while Mouillot and colleagues [36] agreed with Stuart-Smith and colleagues [39] across only 68% of the 2,211 species in common.

Surprisingly, there was also a high heterogeneity in groups with high disagreement (i.e., multiple alternative assignments for species not assigned to the same trophic group). Species classified as “invertivores” according to 1 expert were considered “omnivores,” “piscivores,” or “planktivores” according to other classification schemes (Fig 2). Similarly, species considered “omnivores” by 1 expert were alternatively considered “invertivores,” “herbivores and detritivores,” or “planktivores” by another expert.

Definition of trophic guilds with network analysis

We defined trophic guilds by identifying modules (i.e., combinations of predators and prey) that maximize the weighted modularity of the global network [58]. Our analysis robustly

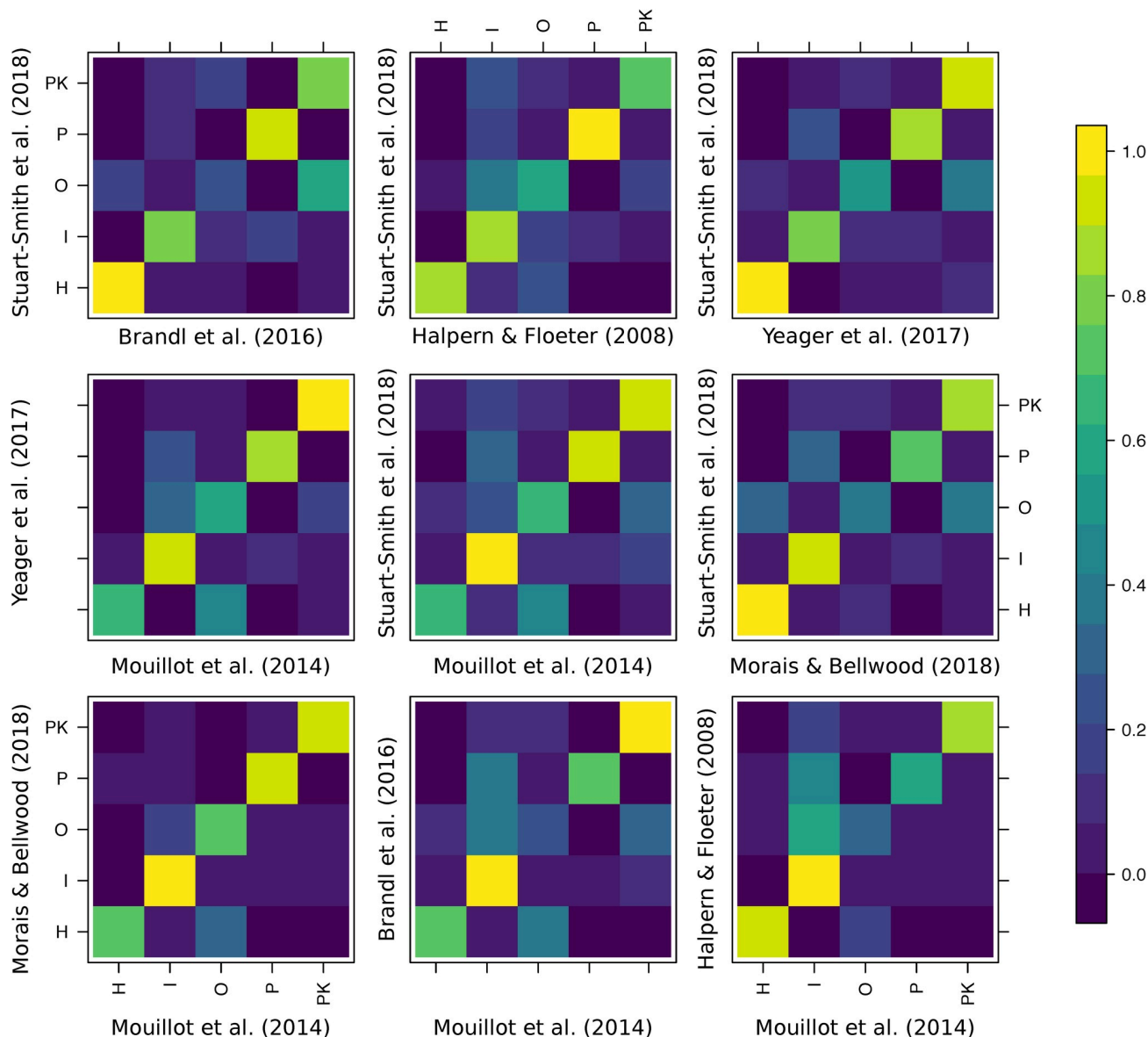


Fig 2. Confusion matrices of the agreement between pairs of experts that share at least 200 species in common and define all 5 trophic categories. Colors represent proportions of species in each trophic guild as classified by experts. The data underlying this figure may be found in https://github.com/valerianoparravicini/Trophic_Fish_2020. H, herbivores and detritivores; I, invertivores; O, omnivores; P, piscivores; PK, planktivores.

<https://doi.org/10.1371/journal.pbio.3000702.g002>

identified 8 distinct modules that correspond to different trophic guilds (Fig 3). We identified these trophic guilds as

1. “Sessile invertivores”: species predominantly feeding on Asteroidea, Bryozoa, Cirripedia, Holothuroidea, Porifera, and Tunicata;
2. “Herbivores, microvores, and detritivores (HMD)”: species primarily feeding on autotrophs, detritus, inorganic material, foraminifera, and phytoplankton;
3. “Corallivores”: species predominantly feeding on Anthozoa and Medusozoa;
4. “Piscivores”: species primarily feeding on Actinopterygii and Cephalopoda;

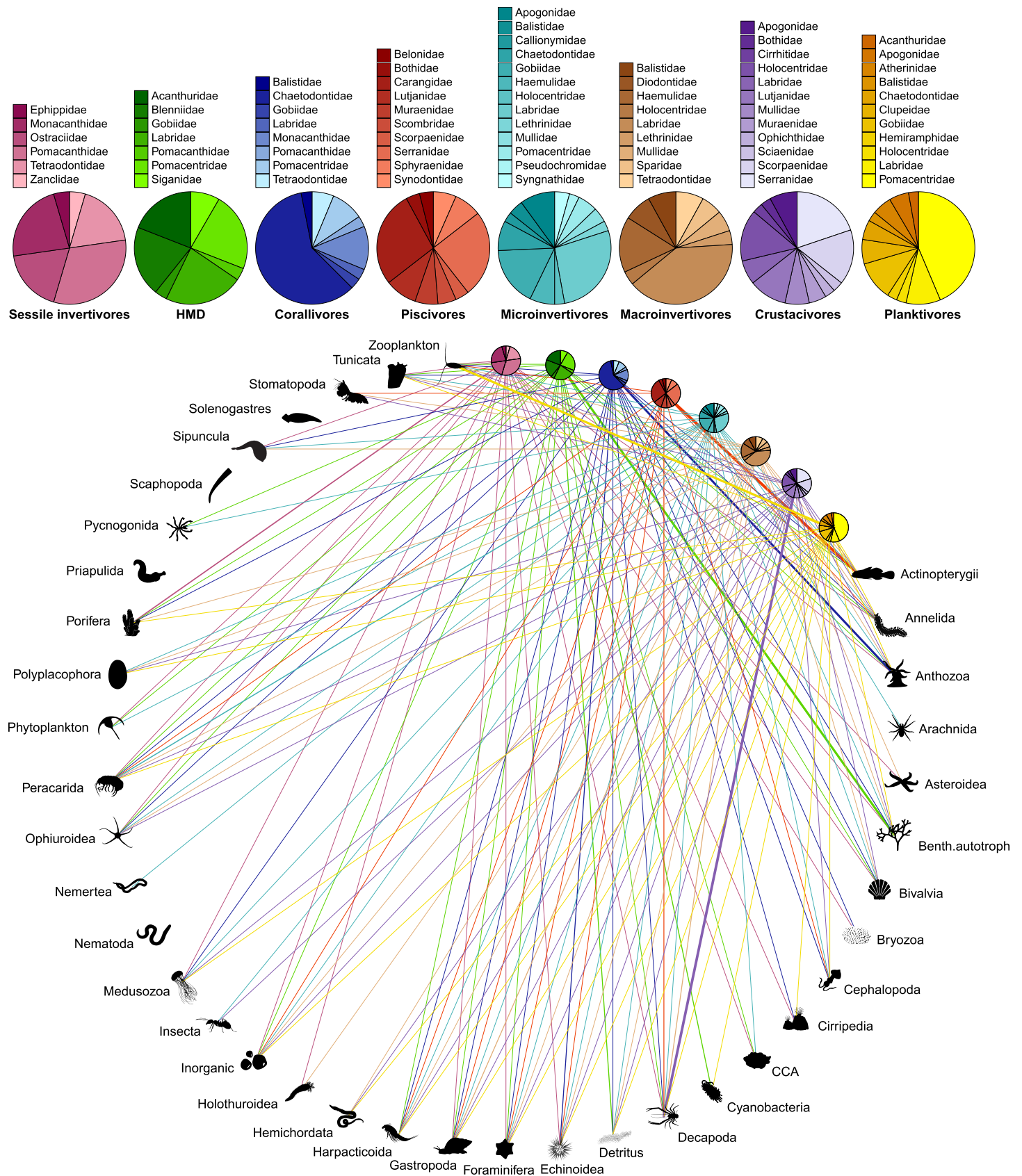


Fig 3. Bipartite network including 615 fish species (grouped into 8 trophic guilds) and their prey items (grouped into 38 categories; see S1 Table). The relative proportion of each prey category consumed by each trophic guild corresponds with the width of each interaction bar. The pie charts show the relative proportion of fish families within each trophic guild. The data underlying this figure may be found in https://github.com/valerianoparravicini/Trophic_Fish_2020. HMD, herbivores, microvores, and detritivores.

<https://doi.org/10.1371/journal.pbio.3000702.g003>

5. “Microinvertivores”: species primarily feeding on Arachnida, Pycnogonida, small Crustacea (Peracarida), and worms (Annelida, Hemichordata, Nematoda, Nemertea, and Sipuncula);
6. “Macroinvertivores”: species primarily feeding on Mollusca (Bivalvia, Gastropoda, Polyplacophora, and Scaphopoda), Echinoidea, and Ophiuroidea;
7. “Crustacivores”: species primarily feeding on large Crustacea (Decapoda and Stomatopoda);
8. “Planktivores”: species mainly feeding on zooplankton, cyanobacteria and Harpacticoida.

Phylogenetic conservatism of trophic guilds

To evaluate the significance of the phylogenetic statistic value ($\delta = 9.37$), we applied a bootstrapping approach and quantified δ after randomly shuffling the trait values 100 times. The median δ of these null models was 0.000199 (95% confidence interval [0.000196, 0.000204]), indicating a strong phylogenetic signal associated with the 8 trophic guilds.

Phylogeny and maximum body size were sufficient to correctly predict the trophic guild of 97% of the species in our dataset. For most families, there was strong phylogenetic conservatism, which resulted in the high confidence of these predictions (Fig 4). Within some families, however, closely related species displayed distinct dietary preferences, as showcased by high negentropy values for families such as Balistidae, Diodontidae, and Labridae.

Given its high predictive performance, we used our Bayesian phylogenetic model to extrapolate the probability of all reef fish species belonging to the 8 trophic guilds and assigned the trophic guild with the highest probability (S3 Table). Using leave-one-out cross validation, the final accuracy of this approach was 65%, which is comparable with other phylogenetically extrapolated traits applications, such as those involving microbial traits [75].

By inspecting the confusion matrix of the leave-one-out cross validation, we obtained more detailed information on the accuracy of the trophic guild predictions (S1 Fig). Most categories were well predicted with our extrapolation approach. In particular, the “sessile invertivores,” “HMD,” and “piscivores” trophic guilds were predicted with high accuracy (77%, 75%, and 73% correct predictions, respectively). The confusion matrix also provided information on incorrectly assigned categories. For example, when “piscivores” were incorrectly assigned, they were mostly classified as “crustacivores.” However, the network plot revealed that the fishes classified as “piscivores” also fed on crustaceans (mostly decapods), so this “incorrect assignment” was grounded in ecological reality and reflected uncertainty within the model. Additionally, the “microinvertivores” trophic guild had the highest proportion of inaccurate predictions (52% correct predictions). Here, species were often misclassified as “crustacivores” or “planktivores.”

Prediction of trophic interactions with machine learning

Using machine learning, our model achieved high predictive performance in quantifying probabilities of pairwise trophic interactions (AUC = 0.99; TSS = 0.94). After calibration with 535 fish species and 3,479 trophic interactions, our model accurately identified 3,410 of these interactions, demonstrating an exceptionally low rate of false negative interactions. In addition, the model accurately predicted absent trophic interactions, with a false positive

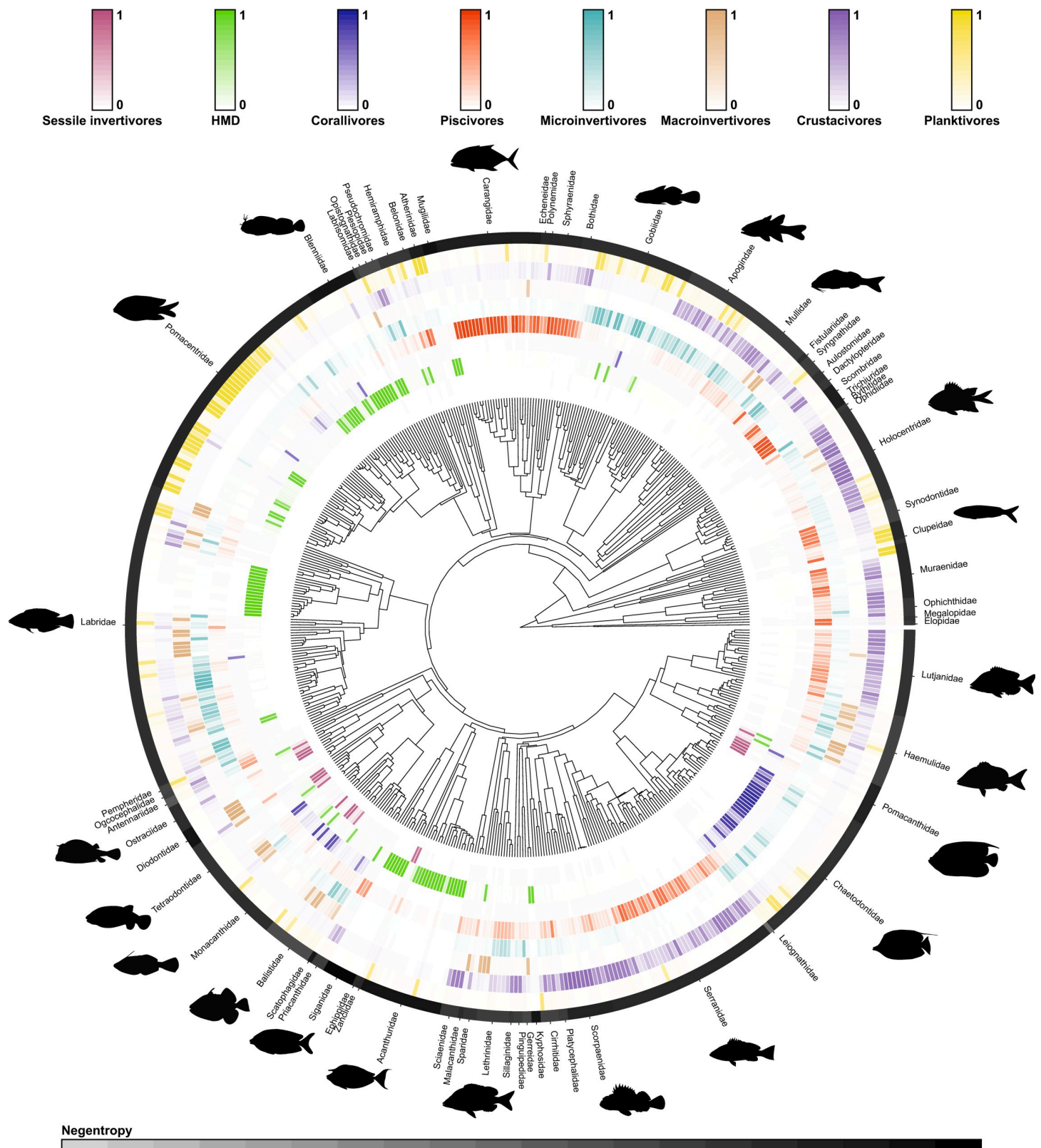


Fig 4. Phylogenetic tree of 535 reef fish species with fitted trophic guild assignments based on empirical dietary data. Trophic guild predictions were made with a Bayesian multinomial phylogenetic regression. The probability of trophic guild assignments for each species is visualized with color scales (depicted above the

phylogenetic tree), with darker colors indicating a higher probability of assignment. In the outer black ring, each distinct segment represents a fish family (with silhouettes included for the most speciose families). Uncertainty of overarching trophic guild assignment for each fish family is visualized with negentropy values (i.e., reverse entropy); thus, darker shades indicate a higher degree of certainty of trophic guild assignment. Fish shapes are available at <https://github.com/simonjbrandl/fishape/tree/master/shapes>. The data underlying this figure may be found in https://github.com/valerianoparravicini/Trophic_Fish_2020.

<https://doi.org/10.1371/journal.pbio.3000702.g004>

interaction rate of only 3.6%. When the model was calibrated with only 5 locations, excluding the data from New Caledonia, the model still performed well (AUC = 0.82; TSS = 0.52). The model correctly detected 82% of the trophic interactions in the New Caledonia dataset, with a false positive interaction rate of 27%. Based on the high predictive performance of the model, we used the full model with all 6 locations to predict the probability of pairwise trophic interactions on a continuous spectrum between the 4,554 reef fish species with available phylogenetic information and our 38 prey categories (S4 Table).

Discussion

Functional ecology requires standardized and reproducible classification schemes to characterize species' niches [76–78]. Rather than relying on expert opinion for the assignment of trophic groups, which often results in variable assignments, we demonstrate that the categorization of discrete trophic guilds and pairwise trophic interactions can be achieved with a quantitative, reproducible framework grounded in empirical data across biogeographic regions. We employed network analysis to partition 535 tropical coral reef fish species into 8 trophic guilds based on a synthesis of globally distributed, community-wide fish dietary analyses, and then we applied a Bayesian phylogenetic model that predicts trophic guilds based on phylogeny and body size, attaining a 5% misclassification error. Moreover, using a machine learning approach, we demonstrate that a continuous spectrum of trophic interactions can also be accurately predicted based on phylogeny and body size. Our framework represents the first implementation of a quantifiable classification scheme for coral reef fishes, which form some of the most diverse vertebrate communities worldwide.

Unlike traditional trophic guilds based on expert opinion [36,37,39,44–49], our trophic approaches are reproducible, provide uncertainty estimates, and can be updated and improved in the future with additional dietary information. In an effort to encourage new, accessible benchmarks to categorize fish trophic guilds, our classification of discrete trophic guilds and probabilities of pairwise trophic interactions are publicly available with this publication. Given the growing number of trait-based studies that assign trophic guilds to understand and monitor ecosystem functioning in our changing world, it is imperative that we establish comparable and reproducible trophic classification frameworks.

Our findings highlight the discordance of expert opinion in the assignment of trophic guilds and the necessity to develop quantifiable and reproducible classification schemes that are accessible to the wider scientific community (c.f. [79]). Despite broad similarities between the trophic guilds reported in the literature and the groups identified by our analysis, our classification scheme reveals a higher level of partitioning among invertebrate-feeding fishes as compared to previously proposed trophic guilds. In the past, invertebrate-feeding fishes were generally considered “sessile invertivores,” “mobile invertivores,” or “omnivores” (e.g., [37,38,48]), but we identify 5 distinct invertebrate-feeding groups: “corallivores,” “sessile invertivores,” “microinvertivores,” “macroinvertivores,” and “crustacivores.” Given the extreme numerical dominance of invertebrates in coral reef environments [80], the collapse of all invertebrate-feeders into 2 or 3 trophic groups was possibly an artefact of expert oversight, and the expansion of invertebrate-feeding trophic guilds to 5 groups stands to improve ecological resolution of fishes feeding on invertebrate prey.

In contrast to the high resolution achieved within invertebrate-feeding groups, our classification achieved limited resolution among the nominally herbivorous species, “HMD.” Across the literature, past classification schemes often separate macroalgal feeders, turf algae croppers, and detritivores (e.g., [36,37,40,41]). The lack of precision in our framework is rooted in the difficulty in distinguishing algae, microbes, and detritus within the alimentary tract of fishes, resulting in the pooling of these ingested items during the visual assessment of fish gut contents. Consequently, species classified as “HMD” may have fundamentally different foraging strategies, dietary preferences, and evolutionary histories [81], which can greatly impact their functional role on coral reefs (e.g., [82]). Thus, while our identified trophic guilds promise increased resolution for fishes that consume animal prey, our identified groupings may not adequately capture consumer–producer dynamics on coral reefs. Emerging techniques, such as gut content metabarcoding, may provide the additional resolution needed to further discriminate prey items in this group [83,84]. Alternatively, coupling diet categorization with other traits, such as feeding behavior, may help to pinpoint the variety of feeding modes that exist within the “HMD” trophic guild.

While our delineation of trophic guilds is applicable to functional studies that employ discrete categories, the continuous output of trophic interaction probabilities holds promise for a variety of other approaches, such as trophic network analyses. On coral reefs, previous studies have employed network analysis to examine human impacts on coral reef food webs [30,33]. However, these studies only incorporate local fish gut content data, which limits their spatial application. Larger-scale network analyses exist (e.g. [85]), but they are predominantly based on co-occurrence patterns and solely consider piscivores, thus neglecting a large portion of marine food webs, which are typically dominated by invertebrate-feeders. Therefore, our demonstrated ability to predict trophic interactions based on phylogeny and body size opens new avenues for marine food web research. Moreover, the high performance of the reduced model to predict pairwise trophic interactions in New Caledonia confirms the potential of our approach to predict probabilities of local trophic interactions for entire food webs.

Our findings add to recent evidence that evolutionary history (i.e., phylogenetic relatedness) is essential to evaluate the ecological traits of fishes (c.f. [86–88]). Recently, taxonomy and body size have been revealed as important predictors of fish diet composition and size structure [89,90], and in the highest resolution analyses of coral reef fish diet, taxonomic family was a better predictor of fish diet than broad trophic guilds [83]. Given the exceedingly low rate of misclassification error in our predictions, we posit that phylogeny is a critical variable that should be consistently considered in the assignment of trophic guilds for reef fishes. Across a plethora of organismal groups (e.g., birds [91], reptiles [92], fishes [93,94], insects [95], parasites [96], and plants [97]), phylogenetic niche conservatism has been alternately supported and dismissed. In our case, when examining fish trophic guilds using 38 prey categories, phylogenetic conservatism is readily apparent at a relatively coarse dietary resolution and may allow us to extrapolate trophic assignments to closely related consumer species and potentially extend this framework to fishes inhabiting other habitats. However, with increasing dietary resolution beyond what is detailed in the present study, phylogenetic signals may weaken [98] since even closely related species may exhibit dietary specialization [83,99]. In the future, with the availability of higher resolution of dietary information, phylogenetic niche conservatism can be easily examined within our framework.

With ongoing environmental and ecological change, a firm grasp on shifts in ecosystem functioning will depend on the reliable assignment of organismal traits [15] and the comparability of trait-based approaches across space, time, and independent studies [77]. Especially in complex, hyperdiverse environments such as coral reefs, it is imperative to standardize how we measure and report these traits to prevent idiosyncratic results based on subjective trait

assignments [27,100]. Trophic guilds are among the most commonly applied trait to assess ecosystem functioning because they directly relate to energy and nutrient fluxes across trophic levels. Thus, our standardized framework represents a major step forward for coral reef functional ecology, while heeding the call for openly accessible, reproducible trait databases [31,78,101]. As trait-based ecology continues to be used to examine disturbances and implement management strategies, our cohesive and accessible framework can provide key insights into the trajectory of coral reef communities.

Further, our results can serve as the foundation for an online platform that permits researchers to collate, update, and utilize trait-based data on coral reef fishes. Similar to current initiatives across the entire tree of life [78], the creation of an online, user-maintained dietary database will facilitate collaboration and traceability in trait-based reef fish research. One challenge will lie in merging visual fish gut content analysis databases with molecular data, such as gut content DNA metabarcoding (e.g., [83]), and biochemical data, such as stable isotope analysis (e.g., [102]), and short-chain fatty acid profiles (e.g., [103]), which indicate nutritional assimilation rather than the simple ingestion of prey items [81]. Despite this challenge, accessibility to a large breadth of reef fish dietary information would improve our framework. Our proposed trophic guilds and probabilities of trophic interactions are model predictions, so they are only as reliable as the underlying dietary data. In addition, these predictions may suffer from extrapolation biases; for example, if limited dietary information exists across species within a taxonomic group, extrapolations to closely related species are more likely to be assigned erroneous trophic guilds. Consequently, an ongoing, extensive compilation of dietary traits across coral reef fishes will continuously improve our predicted trophic guild assignments and pairwise trophic interactions.

Finally, our proposed framework is not limited to coral reef fishes; indeed, trophic guild assignments can be quantifiable, reproducible, and transparent, with the inclusion of uncertainty metrics, across many organismal groups. However, the standardization of trophic guilds is sorely lacking across the ecological literature [79], especially based on quantitative data (e.g., [104]). We posit that a similar approach can be readily applied across a multitude of organisms and environments. In fact, given the paucity of dietary information available for coral reef fishes in comparison to other organisms, particularly birds and mammals, building rigorous, global trophic classification schemes for many other organisms should be readily achievable within our framework. With a quantitative, transparent trophic classification scheme that can be augmented over time and is applicable across ecological systems, our framework represents a significant advancement for trait-based ecology and a viable approach to monitor ecosystem dynamics into the future [78].

Supporting information

S1 Fig. Confusion matrix showcasing the accuracy of the 8 trophic guild predictions from the leave-one-out cross validation based on the extrapolation of the Bayesian phylogenetic model. Trophic guilds include (1) “sessile invertivores,” (2) “herbivores, microvores, and detritivores,” (3) “corallivores,” (4) “piscivores,” (5) “microinvertivores,” (6) macroinvertivores, (7) “crustacivores,” and (8) “planktivores.” The data underlying this figure may be found in https://github.com/valerianoparravicini/Trophic_Fish_2020.
(PNG)

S1 Table. Prey categories used to define the trophic guilds of coral reef fishes.
(CSV)

S2 Table. Summary of the 33 papers used to evaluate expert agreement on reef fish trophic guilds. The column named “Fishes” refers to the number of fish species included in that study. (CSV)

S3 Table. Global extrapolation to infer the probability of each of the 4554 reef fish species to belong to the 8 trophic guilds. The mean and the standard deviation (e.g., p1-8_m, and p1-8_sd) of the posterior probabilities are reported alongside with the mean and standard deviation of the negentropy. (CSV)

S4 Table. Probability of trophic interaction between the 4554 reef fish species and the 38 prey categories according to the extrapolation performed by the machine learning approach. (CSV)

Acknowledgments

We thank the Centre de Recherche Insulaires et Observatoire de l'Environnement (CRIOBE) in Perpignan, France and all the researchers who have made their fish gut content data and/or trait-based fish trophic guild assignments publicly available for use by the wider scientific community

Author Contributions

Conceptualization: Valeriano Parravicini, Jordan M. Casey, Nina M. D. Schittekatte.

Data curation: Valeriano Parravicini, Jordan M. Casey, Nina M. D. Schittekatte, Chloé Pozas-Schacre.

Formal analysis: Valeriano Parravicini, Nina M. D. Schittekatte.

Funding acquisition: Valeriano Parravicini.

Investigation: Valeriano Parravicini, Jordan M. Casey, Nina M. D. Schittekatte, Simon J. Brandl, Chloé Pozas-Schacre, Jérémy Carlot, Graham J. Edgar, Nicholas A. J. Graham, Mireille Harmelin-Vivien, Michel Kulbicki, Giovanni Strona, Rick D. Stuart-Smith.

Methodology: Valeriano Parravicini, Jordan M. Casey, Nina M. D. Schittekatte, Simon J. Brandl, Chloé Pozas-Schacre.

Project administration: Valeriano Parravicini.

Supervision: Valeriano Parravicini, Jordan M. Casey.

Visualization: Valeriano Parravicini, Jordan M. Casey, Nina M. D. Schittekatte, Simon J. Brandl.

Writing – original draft: Valeriano Parravicini, Jordan M. Casey.

Writing – review & editing: Valeriano Parravicini, Jordan M. Casey, Nina M. D. Schittekatte, Simon J. Brandl, Chloé Pozas-Schacre, Jérémy Carlot, Graham J. Edgar, Nicholas A. J. Graham, Mireille Harmelin-Vivien, Michel Kulbicki, Giovanni Strona, Rick D. Stuart-Smith.

References

1. Tilman D, Isbell F, Cowles JM. Biodiversity and Ecosystem Functioning. *Annu Rev Ecol Evol Syst*. 2014; 45:471–93. <https://doi.org/10.1146/annurev-ecolsys-120213-091917>

2. Reiss J, Bridle JR, Montoya JM, Woodward G. Emerging horizons in biodiversity and ecosystem functioning research. *Trends Ecol Evol*. 2009; 24:505–14. <https://doi.org/10.1016/j.tree.2009.03.018> PMID: 19595476
3. Whittaker RH, Levin SA, Root RB. Niche, Habitat, and Ecotope. *Am Nat*. 1973; 107:321–38.
4. Finke DL, Snyder WE. Niche partitioning increases resource exploitation by diverse communities. *Science*. 2008; 321:1488–90. <https://doi.org/10.1126/science.1160854> PMID: 18787167
5. Levine JM, HilleRisLambers J. The importance of niches for the maintenance of species diversity. *Nature*. 2009; 461:254–7. <https://doi.org/10.1038/nature08251> PMID: 19675568
6. Poisot T, Mouquet N, Gravel D. Trophic complementarity drives the biodiversity-ecosystem functioning relationship in food webs. *Ecol Lett*. 2013; 16:853–61. <https://doi.org/10.1111/ele.12118> PMID: 23692591
7. Loreau M, Naeem S, Inchausti P, Bengtsson J, Grime JP, Hector A, et al. Biodiversity and Ecosystem Functioning: Current Knowledge and Future Challenges. *Science*. 2001; 294:804. <https://doi.org/10.1126/science.1064088> PMID: 11679658
8. Terborgh JW. Toward a trophic theory of species diversity. *Proc Natl Acad Sci U S A*. 2015; 112:11415. <https://doi.org/10.1073/pnas.1501070112> PMID: 26374788
9. Moullot D, Bellwood DR, Baraloto C, Chave J, Galzin R, Harmelin-Vivien M, et al. Rare Species Support Vulnerable Functions in High-Diversity Ecosystems. Mace GM, editor. *PLoS Biol*. 2013; 11: e1001569. <https://doi.org/10.1371/journal.pbio.1001569> PMID: 23723735
10. Rosenfeld JS. Functional redundancy in ecology and conservation. *Oikos*. 2002; 98:156–62. <https://doi.org/10.1034/j.1600-0706.2002.980116.x>
11. Wohl DL, Arora S, Gladstone JR. Functional redundancy supports biodiversity and ecosystem function in a closed and constant environment. *Ecology*. 2004; 85:1534–40. <https://doi.org/10.1890/03-3050>
12. Brandl SJ, Bellwood DR. Individual-based analyses reveal limited functional overlap in a coral reef fish community. Hays G, editor. *J Anim Ecol*. 2014; 83:661–70. <https://doi.org/10.1111/1365-2656.12171> PMID: 24164060
13. Wilman H, Belmaker J, Simpson J, de la Rosa C, Rivadeneira MM, Jetz W. EltonTraits 1.0: Species-level foraging attributes of the world's birds and mammals. *Ecology*. 2014; 95:2027–7. <https://doi.org/10.1890/13-1917.1>
14. Kraft NJB, Valencia R, Ackerly DD. Functional traits and niche-based tree community assembly in an Amazonian forest. *Science*. 2008; 322:580. <https://doi.org/10.1126/science.1160662> PMID: 18948539
15. McGill B, Enquist B, Weiher E, Westoby M. Rebuilding community ecology from functional traits. *Trends Ecol Evol*. 2006; 21:178–85. <https://doi.org/10.1016/j.tree.2006.02.002> PMID: 16701083
16. De Deyn GB, Cornelissen JHC, Bardgett RD. Plant functional traits and soil carbon sequestration in contrasting biomes. *Ecol Lett*. 2008; 11:516–31. <https://doi.org/10.1111/j.1461-0248.2008.01164.x> PMID: 18279352
17. Jordano P. Chasing Ecological Interactions. *PLoS Biol*. 2016; 14:e1002559. <https://doi.org/10.1371/journal.pbio.1002559> PMID: 27631692
18. Thompson RM, Brose U, Dunne JA, Hall RO, Hladyz S, Kitching RL, et al. Food webs: reconciling the structure and function of biodiversity. *Trends Ecol Evol*. 2012; 27:689–97. <https://doi.org/10.1016/j.tree.2012.08.005> PMID: 22959162
19. Gravel D, Poisot T, Albouy C, Velez L, Moullot D. Inferring food web structure from predator-prey body size relationships. Freckleton R, editor. *Methods Ecol Evol*. 2013; 4:1083–90. <https://doi.org/10.1111/2041-210X.12103>
20. Sander EL, Wootton JT, Allesina S. Ecological Network Inference From Long-Term Presence-Absence Data. *Sci Rep*. 2017; 7. <https://doi.org/10.1038/s41598-017-07009-x> PMID: 28769079
21. Laigle I, Aubin I, Digel C, Brose U, Boulangeat I, Gravel D. Species traits as drivers of food web structure. *Oikos*. 2018; 127:316–26. <https://doi.org/10.1111/oik.04712>
22. Pichler M, Boreux V, Klein A, Schleuning M, Hartig F. Machine learning algorithms to infer trait-matching and predict species interactions in ecological networks. Carvalho L, editor. *Methods Ecol Evol*. 2020; 11:281–93. <https://doi.org/10.1111/2041-210X.13329>
23. Dalla Riva GV, Stouffer DB. Exploring the evolutionary signature of food webs' backbones using functional traits. *Oikos*. 2016; 125:446–56. <https://doi.org/10.1111/oik.02305>
24. Flynn DFB, Mirotchnick N, Jain M, Palmer MI, Naeem S. Functional and phylogenetic diversity as predictors of biodiversity-ecosystem-function relationships. *Ecology*. 2011; 92:1573–81. <https://doi.org/10.1890/10-1245.1> PMID: 21905424

25. Schmera D, Podani J, Heino J, Erős T, Poff NL. A proposed unified terminology of species traits in stream ecology. *Freshwater Science*. 2015; 34:823–30. <https://doi.org/10.1086/681623>
26. Belmaker J, Parravicini V, Kulbicki M. Ecological traits and environmental affinity explain Red Sea fish introduction into the Mediterranean. *Glob Chang Biol*. 2013; 19:1373–82. <https://doi.org/10.1111/gcb.12132> PMID: 23505033
27. Brandl SJ, Rasher DB, Côté IM, Casey JM, Darling ES, Lefcheck JS, et al. Coral reef ecosystem functioning: eight core processes and the role of biodiversity. *Front Ecol Environ*. 2019; 17:445–54. <https://doi.org/10.1002/fee.2088>
28. Nash KL, Watson RA, Halpern BS, Fulton EA, Blanchard JL. Improving understanding of the functional diversity of fisheries by exploring the influence of global catch reconstruction. *Sci Rep*. 2017; 7:10746. <https://doi.org/10.1038/s41598-017-10723-1> PMID: 28878250
29. Aubin I, Venier L, Pearce J, Moretti M. Can a trait-based multi-taxa approach improve our assessment of forest management impact on biodiversity? *Biodivers Conserv*. 2013; 22:2957–75. <https://doi.org/10.1007/s10531-013-0565-6>
30. Gilarranz LJ, Mora C, Bascompte J. Anthropogenic effects are associated with a lower persistence of marine food webs. *Nat Commun*. 2016; 7. <https://doi.org/10.1038/ncomms10737> PMID: 26867790
31. Madin JS, Anderson KD, Andreasen MH, Bridge TCL, Cairns SD, Connolly SR, et al. The Coral Trait Database, a curated database of trait information for coral species from the global oceans. *Scientific Data*. 2016; 3:160017. <https://doi.org/10.1038/sdata.2016.17> PMID: 27023900
32. Stuart-Smith RD, Bates AE, Lefcheck JS, Duffy JE, Baker SC, Thomson RJ, et al. Integrating abundance and functional traits reveals new global hotspots of fish diversity. *Nature*. 2013; 501:539–42. <https://doi.org/10.1038/nature12529> PMID: 24067714
33. Bascompte J, Melian CJ, Sala E. Interaction strength combinations and the overfishing of a marine food web. *Proc Natl Acad Sci*. 2005; 102:5443–7. <https://doi.org/10.1073/pnas.0501562102> PMID: 15802468
34. McClanahan TR. A coral reef ecosystem-fisheries model: impacts of fishing intensity and catch selection on reef structure and processes. *Ecol Model*. 1995; 80:1–19. [https://doi.org/10.1016/0304-3800\(94\)00042-G](https://doi.org/10.1016/0304-3800(94)00042-G)
35. Ferreira CEL, Floeter SR, Gasparini JL, Ferreira BP, Joyeux JC. Trophic structure patterns of Brazilian reef fishes: a latitudinal comparison. *J Biogeogr*. 2004; 31:1093–106. <https://doi.org/10.1111/j.1365-2699.2004.01044.x>
36. Mouillot D, Villegier S, Parravicini V, Kulbicki M, Arias-Gonzalez JE, Bender M, et al. Functional over-redundancy and high functional vulnerability in global fish faunas on tropical reefs. *Proc Natl Acad Sci*. 2014; 111:13757–62. <https://doi.org/10.1073/pnas.1317625111> PMID: 25225388
37. Parravicini V, Villéger S, McClanahan TR, Arias-González JE, Bellwood DR, Belmaker J, et al. Global mismatch between species richness and vulnerability of reef fish assemblages. Byers JJ, editor. *Ecol Lett*. 2014; 17:1101–10. <https://doi.org/10.1111/ele.12316> PMID: 24985880
38. Micheli F, Mumby PJ, Brumbaugh DR, Broad K, Dahlgren CP, Harborne AR, et al. High vulnerability of ecosystem function and services to diversity loss in Caribbean coral reefs. *Biol Conserv*. 2014; 171:186–94. <https://doi.org/10.1016/j.biocon.2013.12.029>
39. Stuart-Smith RD, Brown CJ, Ceccarelli DM, Edgar GJ. Ecosystem restructuring along the Great Barrier Reef following mass coral bleaching. *Nature*. 2018; 560:92–6. <https://doi.org/10.1038/s41586-018-0359-9> PMID: 30046108
40. Bejarano S, Pardede S, Campbell SJ, Hoey AS, Ferse SCA. Herbivorous fish rise as a destructive fishing practice falls in an Indonesian marine national park. *Ecol Appl*. 2019; 29. <https://doi.org/10.1002/eap.1981> PMID: 31349375
41. Siqueira AC, Bellwood DR, Cowman PF. The evolution of traits and functions in herbivorous coral reef fishes through space and time. *Proc R Soc B Biol Sci*. 2019; 286:20182672. <https://doi.org/10.1098/rspb.2018.2672> PMID: 30963846
42. Mouillot D, Graham NAJ, Villéger S, Mason NWH, Bellwood DR. A functional approach reveals community responses to disturbances. *Trends Ecol Evol*. 2013; 28:167–77. <https://doi.org/10.1016/j.tree.2012.10.004> PMID: 23141923
43. Mason NWH, de Bello F. Functional diversity: a tool for answering challenging ecological questions. *J Veg Sci*. 2013; 24:777–80. <https://doi.org/10.1111/jvs.12097>
44. Brandl SJ, Emslie MJ, Ceccarelli DM, T. Richards Z. Habitat degradation increases functional originality in highly diverse coral reef fish assemblages. *Ecosphere*. 2016; 7. <https://doi.org/10.1002/ecs2.1557>
45. Halpern BS, Floeter SR. Functional diversity responses to changing species richness in reef fish communities. *Mar Ecol Prog Ser*. 2008; 364:147–56.

46. Graham NAJ, Chabanet P, Evans RD, Jennings S, Letourneur Y, Aaron MacNeil M, et al. Extinction vulnerability of coral reef fishes: Reef fish extinction vulnerability. *Ecol Lett*. 2011; 14:341–8. <https://doi.org/10.1111/j.1461-0248.2011.01592.x> PMID: 21320260
47. Morais RA, Bellwood DR. Global drivers of reef fish growth. *Fish Fish*. 2018; 19:874–89. <https://doi.org/10.1111/faf.12297>
48. Yeager LA, Deith MCM, McPherson JM, Williams ID, Baum JK. Scale dependence of environmental controls on the functional diversity of coral reef fish communities. *Glob Ecol Biogeogr*. 2017; 26:1177–89. <https://doi.org/10.1111/geb.12628>
49. Newman MJH, Paredes GA, Sala E, Jackson JBC. Structure of Caribbean coral reef communities across a large gradient of fish biomass. *Ecol Lett*. 2006; 9:1216–27. <https://doi.org/10.1111/j.1461-0248.2006.00976.x> PMID: 17040324
50. Stehman SV. Selecting and interpreting measures of thematic classification accuracy. *Remote Sens Environ*. 1997; 62:77–89. [https://doi.org/10.1016/S0034-4257\(97\)00083-7](https://doi.org/10.1016/S0034-4257(97)00083-7)
51. Hiatt RW, Strasburg DW. Ecological relationships of the fish fauna on coral reefs of the Marshall Islands. *Ecol Monogr*. 1960; 30:65–127. <https://doi.org/10.2307/1942181>
52. Randall JE. Food habits of reef fishes of the West Indies. Hawaii Institute of Marine Biology. 1967:1–94.
53. Hobson ES. Feeding relationships of teleostean fishes on coral reefs in Kona. Hawaii Fishery Bulletin. 1974; 72:915–1031.
54. Harmelin-Vivien M. Ichtyofaune des recifs coralliens de Tulear (Madagascar): ecologie et relations trophiques. L'Université d'Aix-Marseille II. 1979.
55. Sano M, Shimizu M, Nose Y. Food habits of teleostean reef fishes in Okinawa Island, southern Japan. The University of Tokyo; 1984.
56. Boettiger C, Lang DT, Wainwright PC. rfishbase: exploring, manipulating and visualizing FishBase data from R. *J Fish Biol*. 2012; 81:2030–9. <https://doi.org/10.1111/j.1095-8649.2012.03464.x> PMID: 23130696
57. Barrat A, Barthelemy M, Pastor-Satorras R, Vespignani A. The architecture of complex weighted networks. *Proc Natl Acad Sci*. 2004; 101:3747–52. <https://doi.org/10.1073/pnas.0400087101> PMID: 15007165
58. Beckett SJ. Improved community detection in weighted bipartite networks. *R Soc Open Sci*. 2016; 3:140536. <https://doi.org/10.1098/rsos.140536> PMID: 26909160
59. Serafini P. Simulated Annealing for Multi Objective Optimization Problems. In: Tzeng G, Wang H, Wen U, Yu P, editors. Multiple Criteria Decision Making. New York: Springer; 1994. pp. 283–292.
60. Meilă M. Comparing clusterings—an information based distance. *J Multivar Anal*. 2007; 98:873–95. <https://doi.org/10.1016/j.jmva.2006.11.013>
61. Rabosky DL, Chang J, Title PO, Cowman PF, Sallan L, Friedman M, et al. An inverse latitudinal gradient in speciation rate for marine fishes. *Nature*. 2018; 559:392–5. <https://doi.org/10.1038/s41586-018-0273-1> PMID: 29973726
62. Chang J, Rabosky DL, Smith SA, Alfaro ME. An R package and online resource for macroevolutionary studies using the ray-finned fish tree of life. *Methods Ecol Evol*. 2019. <https://doi.org/10.1111/2041-210X.13182>
63. Yu G, Smith DK, Zhu H, Guan Y, Lam TT-Y. ggtree: an R package for visualization and annotation of phylogenetic trees with their covariates and other associated data *Methods in Ecology and Evolution*. 2017; 8:28–36. <https://doi.org/10.1111/2041-210X.12628>
64. Bürkner P-C. brms: An R Package for Bayesian Multilevel Models Using Stan. *J Stat Softw*; Vol 1, Issue 1 (2017). 2017. Available: <https://www.jstatsoft.org/v080/i01>
65. Kembel SW, Cowan PD, Helmus MR, Cornwell WK, Morlon H, Ackerly DD, et al. Picante: R tools for integrating phylogenies and ecology. *Bioinformatics*. 2010; 26:1463–4. <https://doi.org/10.1093/bioinformatics/btq166> PMID: 20395285
66. Garland T, Ives AR. Using the Past to Predict the Present: Confidence Intervals for Regression Equations in Phylogenetic Comparative Methods.: 19.
67. Morales-Castilla I, Matias MG, Gravel D, Araújo MB. Inferring biotic interactions from proxies. *Trends Ecol Evol*. 2015; 30:347–56. <https://doi.org/10.1016/j.tree.2015.03.014> PMID: 25922148
68. Pomeranz JPF, Thompson RM, Poisot T, Harding JS. Inferring predator–prey interactions in food webs. Lecomte N, editor. *Methods in Ecology and Evolution*. 2018 [cited 3 Aug 2020]. <https://doi.org/10.1111/2041-210X.13125>

69. Guénard G, Legendre P, Peres-Neto P. Phylogenetic eigenvector maps: a framework to model and predict species traits. Freckleton R, editor. *Methods Ecol Evol.* 2013; 4:1120–31. <https://doi.org/10.1111/2041-210X.12111>
70. LeDell E, Gill N, Aiello S, Fu A, Candel A, Click C, et al. h2o: R Interface for the “H2O” Scalable Machine Learning Platform. 2020. Available: <https://CRAN.R-project.org/package=h2o>
71. Chen T, Guestrin C. XGBoost: A Scalable Tree Boosting System. *Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining.* 2016; 785–794. <https://doi.org/10.1145/2939672.2939785>
72. Elith J, Leathwick JR, Hastie T. A working guide to boosted regression trees. *J Anim Ecol.* 2008; 77:802–13. <https://doi.org/10.1111/j.1365-2656.2008.01390.x> PMID: 18397250
73. Breiman L. Random forests. *Mach Learn.* 2001; 45:5–32.
74. Allouche O, Tsoar A, Kadmon R. Assessing the accuracy of species distribution models: prevalence, kappa and the true skill statistic (TSS): Assessing the accuracy of distribution models. *J Appl Ecol.* 2006; 43:1223–32. <https://doi.org/10.1111/j.1365-2664.2006.01214.x>
75. Goberna M, Verdú M. Predicting microbial traits with phylogenies. *ISME J.* 2016; 10:959–67. <https://doi.org/10.1038/ismej.2015.171> PMID: 26371406
76. Schneider FD, Fichtmueller D, Gossner MM, Güntsch A, Jochum M, König-Ries B, et al. Towards an ecological trait-data standard. *Methods Ecol Evol.* 2019; 10:2006–19. <https://doi.org/10.1111/2041-210X.13288>
77. Weiss KCB, Ray CA. Unifying functional trait approaches to understand the assemblage of ecological communities: synthesizing taxonomic divides. *Ecography.* 2019; 42:2012–20. <https://doi.org/10.1111/ecog.04387>
78. Gallagher R, Falster DS, Maitner B, Salguero-Gomez R, Vandvik V, Pearse W, et al. The Open Traits Network: Using Open Science principles to accelerate trait-based science across the Tree of Life. <https://doi.org/10.32942/osf.io/kac45>
79. Cano-Barbacid C, Radinger J, García-Berthou E. Reliability analysis of fish traits reveals discrepancies among databases. *Freshwater Biology.* 2020 [cited 11 Feb 2020]. <https://doi.org/10.1111/fwb.13469>
80. Glynn PW, Enochs IC. Invertebrates and Their Roles in Coral Reef Ecosystems. In: Dubinsky Z, Stambler N, editors. *Coral Reefs: An Ecosystem in Transition.* Dordrecht: Springer Netherlands; 2011. pp. 273–325.
81. Clements KD, German DP, Piché J, Tribollet A, Choat JH. Integrating ecological roles and trophic diversification on coral reefs: multiple lines of evidence identify parrotfishes as microphages. *Biol J Linn Soc.* 2016. <https://doi.org/10.1111/bij.12914>
82. Brandl SJ, Bellwood DR. Microtopographic refuges shape consumer-producer dynamics by mediating consumer functional diversity. *Oecologia.* 2016; 182:203–17. <https://doi.org/10.1007/s00442-016-3643-0> PMID: 27147547
83. Casey JM, Meyer CP, Morat F, Brandl SJ, Planes S, Parravicini V. Reconstructing hyperdiverse food webs: Gut content metabarcoding as a tool to disentangle trophic interactions on coral reefs. Mahon A, editor. *Methods Ecol Evol.* 2019; 10:1157–70. <https://doi.org/10.1111/2041-210X.13206>
84. Brandl SJ, Casey JM, Meyer CP. Dietary and habitat niche partitioning in congeneric cryptobenthic reef fish species. *Coral Reefs.* 2020 [cited 28 Jan 2020]. <https://doi.org/10.1007/s00338-020-01892-z>
85. Albouy C, Archambault P, Appeltans W, Araújo MB, Beauchesne D, Cazelles K, et al. The marine fish food web is globally connected. *Nature Ecology & Evolution.* 2019; 3:1153–61. <https://doi.org/10.1038/s41559-019-0950-y> PMID: 31358950
86. Westoby M. Phylogenetic ecology at world scale, a new fusion between ecology and evolution. *Ecology.* 2006; 87:S163–5. [https://doi.org/10.1890/0012-9658\(2006\)87\[163:peawsa\]2.0.co;2](https://doi.org/10.1890/0012-9658(2006)87[163:peawsa]2.0.co;2) PMID: 16922311
87. Floeter SR, Bender MG, Siqueira AC, Cowman PF. Phylogenetic perspectives on reef fish functional traits: Evolution of functional traits. *Biol Rev.* 2018; 93:131–51. <https://doi.org/10.1111/brv.12336> PMID: 28464469
88. Siqueira AC, Morais RA, Bellwood DR, Cowman PF. Trophic innovations fuel reef fish diversification. *Nat Commun.* 2020; 11. <https://doi.org/10.1038/s41467-020-16498-w> PMID: 32472063
89. Soler G, Thomson R, Stuart-Smith R, Smith A, Edgar G. Contributions of body size, habitat and taxonomy to predictions of temperate Australian fish diets. *Mar Ecol Prog Ser.* 2016; 545:239–49. <https://doi.org/10.3354/meps11584>
90. Soler GA, Edgar GJ, Stuart-Smith RD, Smith ADM, Thomson RJ. Predicting the diet of coastal fishes at a continental scale based on taxonomy and body size. *J Exp Mar Biol Ecol.* 2016; 480:1–7. <https://doi.org/10.1016/j.jembe.2016.03.017>

91. Lovette IJ, Hochachka WM. Simultaneous effects of phylogenetic niche conservatism and competition on avian community structure. *Ecology*. 2006; 87:S14–28. [https://doi.org/10.1890/0012-9658\(2006\)87\[14:seopnc\]2.0.co;2](https://doi.org/10.1890/0012-9658(2006)87[14:seopnc]2.0.co;2) PMID: 16922299
92. Knouft JH, Losos JB, Glor RE, Kolbe JJ. Phylogenetic analysis of the evolution of the niche in lizards of the *Anolis sagrei* group. *Ecology*. 2006; 87:S29–38. [https://doi.org/10.1890/0012-9658\(2006\)87\[29:paoteo\]2.0.co;2](https://doi.org/10.1890/0012-9658(2006)87[29:paoteo]2.0.co;2) PMID: 16922300
93. Peres-Neto PR. Patterns in the co-occurrence of fish species in streams: the role of site suitability, morphology and phylogeny versus species interactions. *Oecologia*. 2004; 140:352–60. <https://doi.org/10.1007/s00442-004-1578-3> PMID: 15138880
94. Gajdzik L, Aguilar-Medrano R, Frédérich B. Diversification and functional evolution of reef fish feeding guilds. *Ecol Lett*. 2019. <https://doi.org/10.1111/ele.13219> PMID: 30648337
95. Weiblen GD, Webb CO, Novotny V, Basset Y, Miller SE. Phylogenetic dispersion of host use in a tropical insect herbivore community. *Ecology*. 2006; 87:S62–75. [https://doi.org/10.1890/0012-9658\(2006\)87\[62:pdohui\]2.0.co;2](https://doi.org/10.1890/0012-9658(2006)87[62:pdohui]2.0.co;2) PMID: 16922303
96. Mouillot D, R. Krasnov B, I. Shenbrot G, J. Gaston K, Poulin R. Conservatism of host specificity in parasites. *Ecography*. 2006; 29:596–602. <https://doi.org/10.1111/j.0906-7590.2006.04507.x>
97. Silvertown J, Dodd M, Gowing D, Lawson C, McConway K. Phylogeny and the hierarchical organization of plant diversity. *Ecology*. 2006; 87:S39–49. [https://doi.org/10.1890/0012-9658\(2006\)87\[39:pathoo\]2.0.co;2](https://doi.org/10.1890/0012-9658(2006)87[39:pathoo]2.0.co;2) PMID: 16922301
98. Mazel F, Pennell MW, Cadotte MW, Diaz S, Dalla Riva GV, Grenyer R, et al. Prioritizing phylogenetic diversity captures functional diversity unreliably. *Nat Commun*. 2018; 9. <https://doi.org/10.1038/s41467-018-05126-3> PMID: 30038259
99. Leray M, Alldredge AL, Yang JY, Meyer CP, Holbrook SJ, Schmitt RJ, et al. Dietary partitioning promotes the coexistence of planktivorous species on coral reefs. *Mol Ecol*. 2019; 28:2694–710. <https://doi.org/10.1111/mec.15090> PMID: 30933383
100. Bellwood DR, Streit RP, Brandl SJ, Tebbett SB. The meaning of the term ‘function’ in ecology: A coral reef perspective. *Funct Ecol*. 2019; 33:948–61. <https://doi.org/10.1111/1365-2435.13265>
101. Jeliaskov A, Mijatovic D, Chantepie S, Andrew N, Arlettaz R, Barbaro L, et al. A global database for metacommunity ecology, integrating species, traits, environment and space. *Scientific Data*. 2020; 7. <https://doi.org/10.1038/s41597-019-0344-7> PMID: 31913312
102. Eurich J, McCormick M, Jones G. Habitat selection and aggression as determinants of fine-scale partitioning of coral reef zones in a guild of territorial damselfishes. *Mar Ecol Prog Ser*. 2018; 587:201–15. <https://doi.org/10.3354/meps12458>
103. Choat HJ, Clements KD, Robbins WD. The trophic status of herbivorous fishes on coral reefs: I. Dietary analyses *Marine Biology*. 2002; 140:613–23. <https://doi.org/10.2527/2002.80102726x> PMID: 12413096
104. González-Salazar C, Martínez-Meyer E, López-Santiago G. A hierarchical classification of trophic guilds for North American birds and mammals. *Revista Mexicana de Biodiversidad*. 2014; 85:931–41. <https://doi.org/10.7550/rmb.38023>